Alpine Entomology 6 2022, 1–6 | DOI 10.3897/alpento.6.80405

PENSOFT

Alpine Entomology

Lack of genetic structure suggests high connectivity of *Parnassius phoebus* between nearby valleys in the Alps

Andreas Jaun[1], Hans-Peter Wymann[1,2], Kay Lucek[3]

1 *InfoNatura, Niederliweg 10, 3700 Spiez, Switzerland*
2 *Natural History Museum Bern, Bernastrasse 15, 3005 Bern, Switzerland*
3 *Department of Environmental Sciences, University of Basel, Schönbeinstrasse 6, 4056 Basel, Switzerland*

http://zoobank.org/C8A457C7-7896-4EF1-96C6-BF462C7D8219

Corresponding author: Kay Lucek (kay.lucek@unibas.ch)

Academic editor: Peter Huemer ♦ **Received** 11 January 2022 ♦ **Accepted** 13 March 2022 ♦ **Published** 24 March 2022

Abstract

The spatial scale of intraspecific genetic connectivity and population structure are important aspects of conservation genetics. However, for many species these properties are unknown. Here we used genomic data to assess the genetic structure of the small Apollo butterfly (*Parnassius phoebus* Fabricius, 1793; Lepidoptera: Papilionidae) across three nearby valleys in the Central Swiss Alps. One of the valleys is currently used for hydropower production with future plans to raise the existing dam wall further. We found no significant genetic structure, suggesting a currently high connectivity of this species in our studied region.

Key Words

Lepidoptera, conservation, gene flow, genomics, *Parnassius sacerdos*

Introduction

The maintenance of genetic diversity is a key target of current conservation efforts because such diversity is thought to enable species to cope with changing environments (DeWoody et al. 2021). Among the factors that can reduce genetic diversity are habitat fragmentation and global climate change (Pauls et al. 2013; Schlaepfer et al. 2018). Alpine environments may especially be threatened by climate change (Engler et al. 2011), the latter often promoting the subdivision of locally adapted species (Jordan et al. 2016). The scale at which intraspecific gene flow occurs is thus an important property of a species with significant implications for conservation and management. However, the spatial scale of genetic connectivity is often unknown as its assessment either requires large-scale mark-recapture studies or genomic data (Gagnaire et al. 2015).

Here, we assessed the potential for intraspecific gene flow in an alpine butterfly – the small Apollo (*Parnassius phoebus* Fabricius, 1793; Lepidoptera: Papilionidae). The species occurs locally in alpine environments from Alaska over Russia to the Alps (Todisco et al. 2012). Many of its allopatric populations have been described as distinct subspecies whose taxonomic status has though remained elusive (Weiss and Rigout 2005). For example, there is an ongoing debate if *P. phoebus* from the Alps should be named *P. sacerdos* (International Commission on Zoological Nomenclature 2017) or not (Bálint 2021), where *P. sacerdos* and Eurasian *P. phoebus* are polyphyletic based on mitochondrial haplotypes (Todisco et al. 2012). Given the unresolved taxonomy, we use *P. phoebus* here, which is consistent with the current Swiss red list for butterflies (Wermeille et al. 2014). *P. phoebus* subspecies differ often phenotypically from each other but intraspecific phenotypic variation also occurs at smaller scales. Indeed, a former study on alpine melanism, highlighted the adaptive value of increased melanism with increased elevation and latitude in *P. phoebus* (Guppy 1986). Males that were darker on their hindwings spent a greater proportion of time in flight at low air temperatures and showed increased movement (Guppy

Copyright *Andreas Jaun et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1986). Importantly, the global diversity within *P. phoebus* is young, *i.e.*, evolved over the last ~125'000 years, where geographically distant populations within a continent diverged as recently as 10'000–50'000 years ago (Todisco et al. 2012). Like for other species of this genus, *P. phoebus* is thought to have moderate dispersal capabilities, being able to fly from some hundred metres to few kilometres (Guppy 1986; Brommer and Fred 1999). However, natural barriers has been shown to limit intraspecific gene flow in other *Parnassius* species (Keyghobadi et al. 1999), but to which degree this is true for *P. phoebus* is not known.

Parnassius phoebus is a univoltine species and in the Alps can be found in humid, often flooded habitats with mostly extensive stands of *Saxifraga aizoides*, the primary larval food plant of this species (Lepidopterologen Arbeitsgruppe 1987). Habitats include relatively flat headwaters and riparian zones of small and large watercourses, often with alluvial plains in the subalpine and alpine and occasionally the montane zones. In the Bernese Alps, the species can be found from 1400 to 2300 m elevation, occurring both on limestone and silicate rock substrates. Imagoes feed on nectar from a range of plants, including thistles, *Origanum* and various cushion-forming plants, such as *Saxifraga*. Eggs are generally not directly laid on the host plant, but either on dried plants in its vicinity or directly on the soil substrate (Lepidopterologen Arbeitsgruppe 1987).

We used nuclear genomic data to assess the potential for gene flow among individuals collected from three nearby valleys in the Central Swiss Alps (Fig. 1). We focused on this region because the Trift valley experienced significant past and future anthropogenic alterations as a consequence of artificial damming for hydropower production (Haeberli et al. 2016; Guillén-Ludeña et al. 2018). This, together with the impact of climate change could thus render *P. phoebus* locally vulnerable, especially if current intraspecific gene flow would be limited (Condamine and Sperling 2018).

Methods

Sampling

We collected a total of 18 butterflies during summers 2015–2020. Sampling was conducted in three valleys in the Central Swiss Alps (Susten (N=6), Trift (N=8), Wenden (N=4), Fig. 1, Suppl. material 1: Table S1). We captured all individuals with hand nets and killed them with an overdose of ethyl acetate. Full bodies were dried for further genetic analyses.

Genetic data processing

We genotyped all individuals using single-end restriction-site associated DNA (RAD) sequencing with the restriction enzyme *Pst*I. For all individuals we extracted the DNA from thorax tissue using the Qiagen DNeasy Blood and Tissue kit (Qiagen, Zug, Switzerland) following the manufacturer's protocol. Library preparation and sequencing on one Illumina HiSeq 4000 lane was outsourced to Floragenex (Portland, OR, USA). All genomic data is archived on NCBI (BioProject ID: PRJNA814465).

We filtered all obtained genomic data following (Lucek et al. 2020), *i.e.*, we only retained reads with an intact *Pst*I restriction site, followed by de-multiplexing and barcode-trimming with *process_radtags* from Stacks 1.48 (Catchen et al. 2013). Using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), we then removed reads containing bases with a Phred quality score <10 or more than 5% of base pairs with quality <30. This approach yielded ~18.5 million high quality reads in total for our analysis. Given the lack of a *Phoebus* reference genome, we generated a *de novo* assembly of RAD-tags using all filtered reads for all individuals with *ustacks* 1.48 (Catchen et al. 2013) with the following settings: minimum stack size of 50 reads, a maximum of three base pairs of difference for stacks to be merged, excluding loci with unusually high coverage to avoid repetitive regions. The initial *de novo* assembly consisted of 11'004 contigs. To further identify and remove exogenous contigs from the assembly, we compared the assembly against the NCBI GenBank nucleotide collection with the *blastn* function from BLAST+ 2.7.1 (Camacho et al. 2009). A total of 40 or 0.4% of all contigs were of exogenous origin and we removed them from the initial assembly.

In a next step, we mapped the reads of each individual against our reference assembly with minimap2 2.2 (Li 2018) and genotyped all specimens with BCFtools 1.10.2 (Danecek and McCarthy 2017). We filtered the genotypes with VCFtools 0.1.16 (Danecek et al. 2011) to remove indels, to include only bi-allelic polymorphic sites with a minimal depth of six and a minimal genotype quality of 20, employing a minor allele frequency filter of 0.03 and allowing up to 50% of missing data per site. Due to high rates of missing data, two specimens were filtered out (Suppl. material 1: Table S1). The overall filtering resulted in 5157 SNP sites available for our downstream analyses.

Genetic analyses

To test for an individual based genetic structure, we first employed a phylogenomic analysis comprising all retained specimens. We used RAXML 8.2.11 (Stamatakis 2014) implementing a generalised time-reversible (GTR) model with optimised substitution rates and a gamma model of rate heterogeneity. We further applied an ascertainment bias correction to account for the fact that we only used polymorphic SNP positions with the ASC_GTRGAMMA function implemented in RAXML. Significance was assessed using 1000 bootstrap replicates followed by a thorough maximum likelihood search.

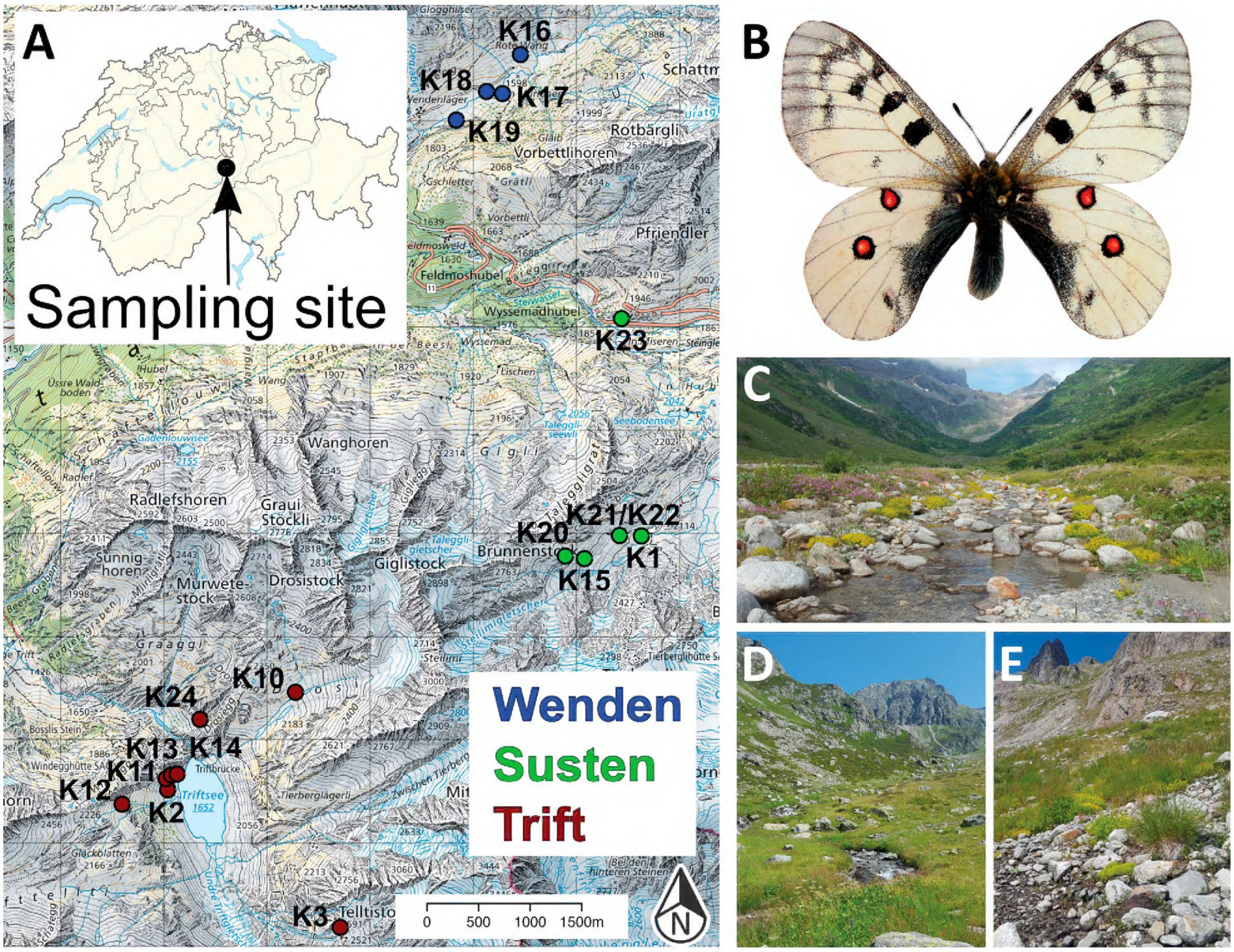

Figure 1. Overview of our sampled sites. **A.** Map depicting the sampling locations of all collected individuals from the central Swiss Alps with the inset depicting the sampling site in Switzerland (see Table S1 for details). Circle colour indicates the different valleys. For each individual the respective sample ID is given (see Table S1). Map source: Federal Office of Topography swisstopo; **B.** Example of *Parnassius phoebus* (individual K13); **C–E.** Habitat pictures for Wenden, Trift and Susten, respectively.

We next inferred population structure with Admixture 1.3.0, which implements a likelihood approach to estimate ancestry (Alexander et al. 2009). We ran ADMIXTURE by varying the number of assumed populations, *i.e.*, K, from 1 to 5 and performed a cross-validation test to determine the optimal value of K. In a second step we used a principal component (PC) analysis as implemented in GenoDive 3.0.5 (Meirmans 2020) to visualize the genetic relationship among individuals.

Finally, we estimated the overall level of pairwise genetic differentiation (F_{ST}) among individuals from the three valleys (Susten, Trift, Wenden; see Suppl. material 1: Table S1) using GenoDive, with 1000 bootstrap iterations to estimate significance. Because genetic differentiation would only occur at few loci that experience direct or indirect selection in the case of recent divergence (Seehausen et al. 2014), we also performed locus-by-locus F_{ST} in Genodive analyses between Trift individuals and individuals from Susten and Wenden combined.

Results

The bootstrap approach employed in our RAXML analysis found no significant node splits (*i.e.* >95% bootstrap support), suggesting the absence of a detectable differentiation among individuals. Similarly, no clustering occurred related to the three different valleys (Fig. 2a). The best number of genetic clusters as inferred by Admixture was likewise one (K=1), where the subsequent model assuming two genetic cluster showed no clustering by valleys (Fig. 2c). The two leading PC axes accounted for 9.3 and 8.4% of the total variation respectively and only here some individuals from the Trift valley seemed to be differentiated from the other individuals along PC1 (Fig. 2b).

The level of pairwise genetic differentiation among valleys was generally low and non-significant (Susten vs. Trift: F_{ST} = 0.005, p = 0.195; Susten vs. Wenden: F_{ST} = 0.017, p = 0.143; Trift vs. Wenden: F_{ST} = 0.021, p = 0.177). This was similarly true when individuals from Susten and Wenden were pooled (F_{ST} = 0.006, p = 0.059; Fig. 3). The locus-by-locus analysis for the

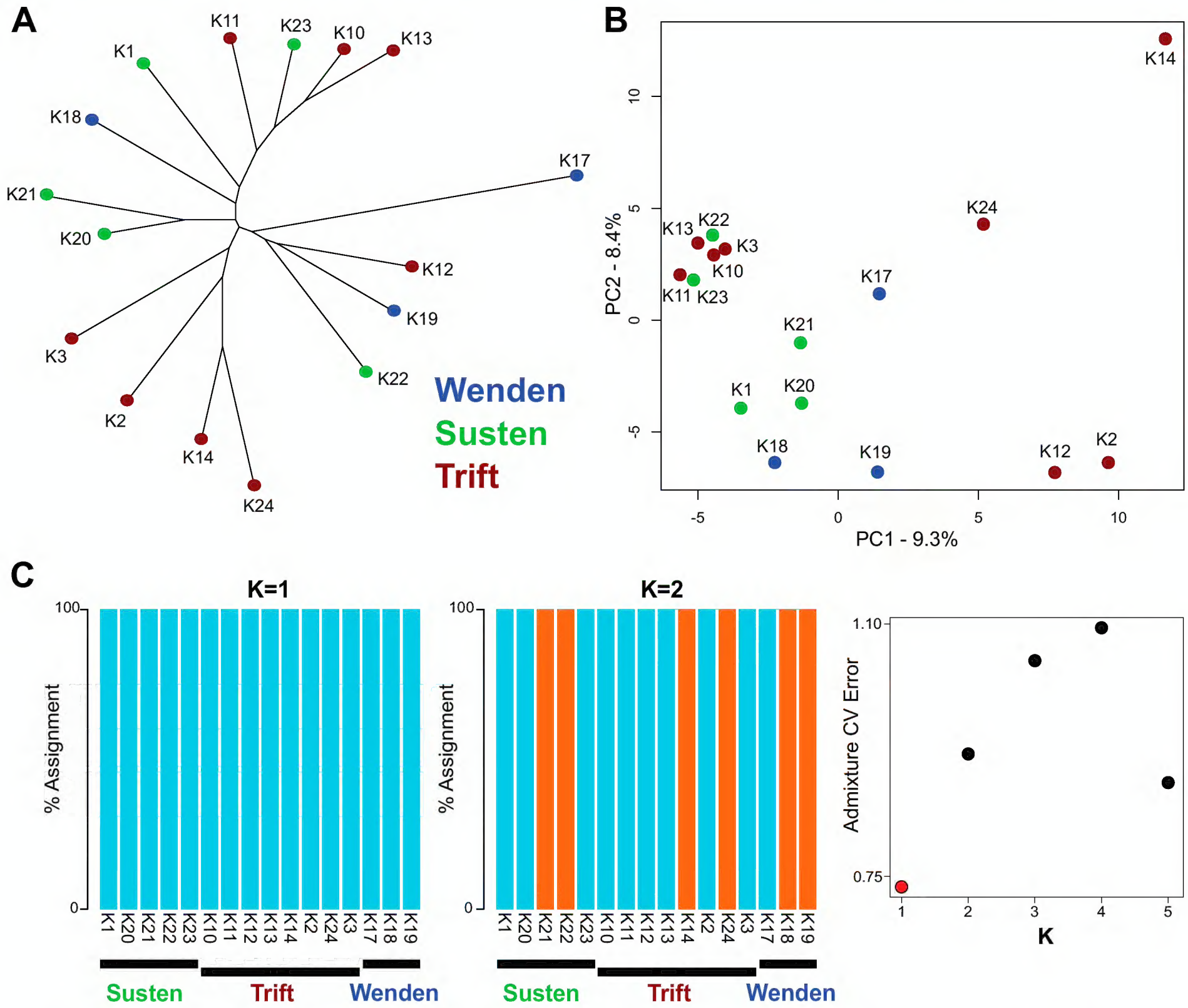

Figure 2. Summary of the individual based genetic analyses. **A.** Unrooted phylogram based on a RAxML analysis. None of the nodes were statistically significant (all with <95% bootstrap support); **B.** Scores of a principal component (PC) analysis for the two leading axes. C – Individual based assignments as inferred by Admixture. One genetic cluster (K=1) was the best supported number as inferred by the cross validation (CV) error. For **A, B** colours depict valleys. For each individual the respective sample ID is given (see Fig. 1, Suppl. material 1: Table S1).

same comparison identified only 11 SNPs with an F_{ST} > 0.20 (Fig. 3), however, none of the associated contigs could be mapped to a known gene by BLAST.

Discussion

Using genomic data, we found a lack of genetic structure among individuals of the small Apollo *Parnassius phoebus* that could be attributed to the three valleys in close proximity, *i.e.*, being 4–8 km apart, which we sampled in the central Swiss Alps (Figs 1, 2). Our results thus suggest a high connectivity of this species in our studied region. Consequently, the valleys, mountain ridges, glaciers or other potentially unsuitable habitat structures in our studied region (Fig. 1) do not present strong barriers to gene flow. This finding contrasts with observations in other *Parnassius* species where such geographic structures resulted in fine-scale population structure (Brommer and Fred 1999; Keyghobadi et al. 1999). While the absence of significant genetic differentiation, as estimated by F_{ST}, may also highlight the statistical limitations given the sample size of our study, the individual-based analyses that we applied would allow to detect potential fine-scale structure (Rieder et al. 2019).

P. phoebus is an evolutionary young species that has moreover recolonized the studied area only after the last glaciation period (Todisco et al. 2012). Consequently, even if local adaptation would have occurred, the respective populations may not necessarily have had enough time to accumulate genetic differentiation beyond few genomic regions that experience selection (Nosil 2012; Seehausen et al. 2014). Indeed, our locus-based analysis identified

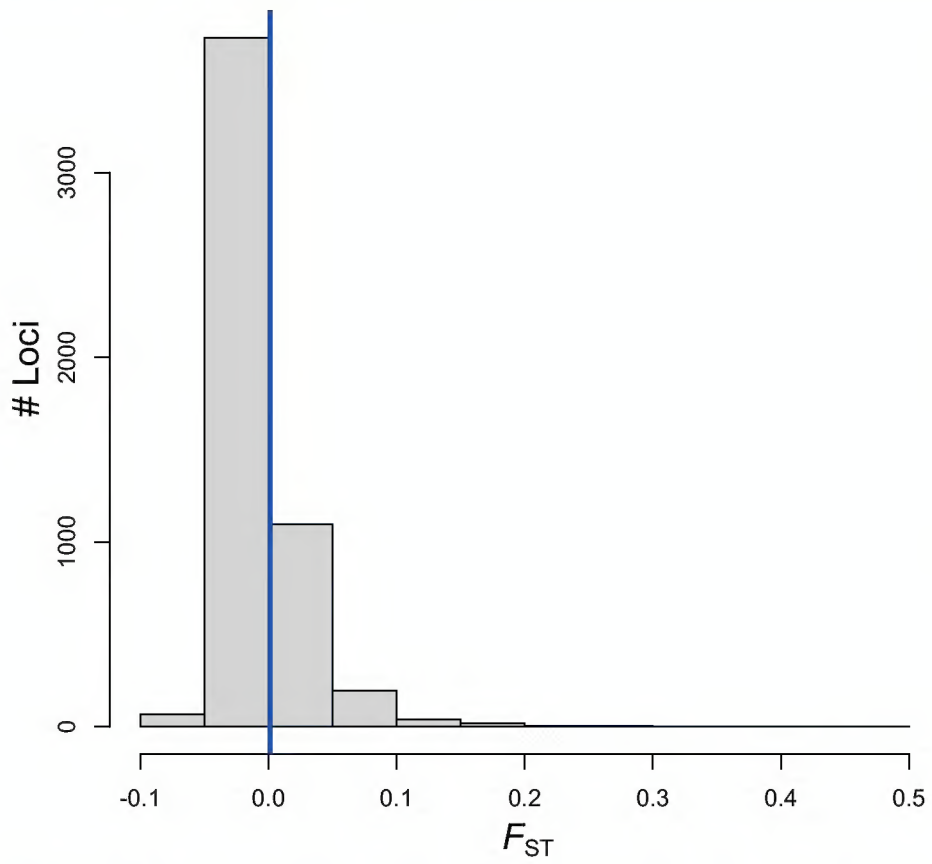

Figure 3. Histogram of the genetic differentiation (F_{ST}) between individuals from the Trift valley and the two other valleys combined, calculated for each locus separately. The blue line indicated the global $F_{ST} = 0.0057$.

very few sites of accentuated differentiation (Fig. 3). Such genomic differentiation at only few target loci may be consistent with a potential very early stage of divergence-with-gene-flow, where further differentiation depends on the evolution of barriers to gene flow (Nosil 2012). However, the interpretation of such genomic regions has to be done with care, as they can also emerge through non-adaptive processes including genetic drift (Ravinet et al. 2017). Lastly, both the lack of significant genomic differentiation and the limited number of loci that showed accentuated differentiation could reflect a limited resolution given the restricted number of polymorphic SNPs available for our analyses and the absence of a reference genome.

A high connectivity despite potential natural barriers may suggest that *P. phoebus* could be less affected by future anthropogenic modifications in the studied area (Haeberli et al. 2016; Guillén-Ludeña et al. 2018). However, such modifications will act combined with the effects of climate change, which is thought to be a main threat for species of the genus *Parnassius* (Condamine and Sperling 2018). Although *P. phoebus* can likely track its climatic niche by shifting its range up the mountains until they can go no higher, the species also depends on the availability of host plants, which can be equally affected by both factors (Condamine and Sperling 2018). Therefore, from a conservation perspective, it would be advisable to broaden the geographic scope of our study to identify the scale of potential population structure in *P. phoebus* across the Alps, ideally with denser genomic data. In addition, future anthropogenic habitat modifications, as it is planned for the Trift valley (Ehrbar et al. 2018), should be accompanied by a genetic monitoring for both *P. phoebus* and its host plant.

Acknowledgements

Fieldwork and sequencing were funded by the Kraftwerke Oberhasli (KWO) AG. We would especially thank Magdalena Rohrer from the KWO for her support throughout the project. We further thank Sebastian Wymann for his outstanding efforts during fieldwork. We thank Peter Huemer, Thomas Schmitt and two anonymous reviewers for their insightful comments that helped us improve the manuscript. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19: 1655–1664. https://doi.org/10.1101/gr.094052.109

Bálint Z (2021) Comment (Case 3767) – More support for the proposed conservation of prevailing usage of the specific name *Papilio phoebus* Fabricius, 1793 (currently *Parnassius phoebus*; Insecta, Lepidoptera), and that of *Doritis ariadne* Lederer, 1853 (currently *Parnassius ariadne*) by the designation of a neotype. The Bulletin of Zoological Nomenclature 78: 38–41. https://doi.org/10.21805/bzn.v78.a011

Brommer Jon E, Fred MS (1999) Movement of the Apollo butterfly *Parnassius apollo* related to host plant and nectar plant patches. Ecological Entomology 24: 125–131. https://doi.org/10.1046/j.1365-2311.1999.00190.x

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinformatics 10: e421. https://doi.org/10.1186/1471-2105-10-421

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Molecular Ecology 22: 3124–3140. https://doi.org/10.1111/mec.12354

Condamine F, Sperling F (2018) Anthropogenic threats to high-altitude parnassian diversity. News of the Lepidopterists' Society 60: 94–99.

Danecek P, McCarthy SA (2017) BCFtools/csq: haplotype-aware variant consequences. Bioinformatics 33: 2037–2039. https://doi.org/10.1093/bioinformatics/btx100

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. Bioinformatics 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

DeWoody JA, Harder AM, Mathur S, Willoughby JR (2021) The long-standing significance of genetic diversity in conservation. Molecular Ecology 30: 4147–4154. https://doi.org/10.1111/mec.16051

Ehrbar D, Schmocker L, Vetsch D, Boes R (2018) Hydropower potential in the periglacial environment of Switzerland under climate change. Sustainability 10: e2794. https://doi.org/10.3390/su10082794

Engler R, Randin CF, Thuiller W, Dullinger S, Zimmermann NE, Araújo MB, Pearman PB, Le Lay G, Piedallu C, Albert CH, Choler P, Coldea G, De Lamo X, Dirnböck T, Gégout J-C, Gómez-García D, Grytnes J-A, Heegaard E, Høistad F, Nogués-Bravo D, Normand S, Puşcaş M, Sebastià M-T, Stanisci A, Theurillat J-P, Trivedi MR,

Vittoz P, Guisan A (2011) 21st century climate change threatens mountain flora unequally across Europe. Global Change Biology 17: 2330–2341. https://doi.org/10.1111/j.1365-2486.2010.02393.x

Gagnaire P, Broquet T, Aurelle D, Viard F, Souissi A, Bonhomme F, Arnaud-Haond S, Bierne N (2015) Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. Evolutionary Applications 8: 769–786. https://doi.org/10.1111/eva.12288

Guillén-Ludeña S, Manso P, Schleiss A (2018) Multidecadal sediment balance modelling of a cascade of Alpine reservoirs and perspectives based on climate warming. Water 10: e1759. https://doi.org/10.3390/w10121759

Guppy CS (1986) The adaptive significance of alpine melanism in the butterfly *Parnassius phoebus* F. (Lepidoptera: Papilionidae). Oecologia 70: 205–213. https://doi.org/10.1007/BF00379241

Haeberli W, Buetler M, Huggel C, Friedli TL, Schaub Y, Schleiss AJ (2016) New lakes in deglaciating high-mountain regions – opportunities and risks. Climatic Change 139: 201–214. https://doi.org/10.1007/s10584-016-1771-5

International Commission on Zoological Nomenclature (2017) Opinion 2382 (Case 3637) – Conservation of the accustomed usage of *Papilio phoebus* De Prunner, 1798 by suppression of *Papilio phoebus* Fabricius, 1793 not approved (Insecta, Lepidoptera, papilionidae). The Bulletin of Zoological Nomenclature 73: 148–149. https://doi.org/10.21805/bzn.v73i2.a21

Jordan S, Giersch JJ, Muhlfeld CC, Hotaling S, Fanning L, Tappenbeck TH, Luikart G (2016) Loss of genetic diversity and increased subdivision in an endemic Alpine stonefly threatened by climate change. PLoS ONE 11: e0157386. https://doi.org/10.1371/journal.pone.0157386

Keyghobadi N, Roland J, Strobeck C (1999) Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). Molecular Ecology 8: 1481–1495. https://doi.org/10.1046/j.1365-294x.1999.00726.x

Lepidopterologen Arbeitsgruppe (1987) Tagfalter und ihre Lebensräume. Schweiz und angrenzende Gebiete. Arten, Gefährdung, Schutz. Schweizerischer Bund für Naturschutz, Basel, 516 pp.

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34: 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Lucek K, Butlin RK, Patsiou T (2020) Secondary contact zones of closely-related *Erebia* butterflies overlap with narrow phenotypic and parasitic clines. Journal of Evolutionary Biology 33: 1152–1163. https://doi.org/10.1111/jeb.13669

Meirmans PG (2020) Genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. Molecular Ecology Resources 20: 1126–1131. https://doi.org/10.1111/1755-0998.13145

Nosil P (2012) Ecological speciation. Oxford University Press, Oxford, UK, 271 pp. https://doi.org/10.1093/acprof:osobl/9780199587100.001.0001

Pauls SU, Nowak C, Bálint M, Pfenninger M (2013) The impact of global climate change on genetic diversity within populations and species. Molecular Ecology 22: 925–946. https://doi.org/10.1111/mec.12152

Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, Noor MAF, Mehlig B, Westram AM (2017) Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. Journal of Evolutionary Biology 30: 1450–1477. https://doi.org/10.1111/jeb.13047

Rieder JM, Vonlanthen P, Seehausen O, Lucek K (2019) Allopatric and sympatric diversification within roach (*Rutilus rutilus*) of large pre-alpine lakes. Journal of Evolutionary Biology 32: 1174–1185. https://doi.org/10.1111/jeb.13502

Schlaepfer DR, Braschler B, Rusterholz H-P, Baur B (2018) Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. Ecosphere 9: e02488. https://doi.org/10.1002/ecs2.2488

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G-P, Bank C, Brännström Å, Brelsford A, Clarkson CS, Eroukhmanoff F, Feder JL, Fischer MC, Foote AD, Franchini P, Jiggins CD, Jones FC, Lindholm AK, Lucek K, Maan ME, Marques DA, Martin SH, Matthews B, Meier JI, Möst M, Nachman MW, Nonaka E, Rennison DJ, Schwarzer J, Watson ET, Westram AM, Widmer A (2014) Genomics and the origin of species. Nature Reviews Genetics 15: 176–192. https://doi.org/10.1038/nrg3644

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Todisco V, Gratton P, Zakharov EV, Wheat CW, Sbordoni V, Sperling FAH (2012) Mitochondrial phylogeography of the Holarctic *Parnassius phoebus* complex supports a recent refugial model for alpine butterflies. Journal of Biogeography 39: 1058–1072. https://doi.org/10.1111/j.1365-2699.2011.02675.x

Weiss J, Rigout J (2005) The Parnassiinae of the world. Part 4. Sciences Nat, Venette, France.

Wermeille E, Chittaro Y, Gonseth Y (2014) Rote Liste der Tagfalter und Widderchen: Papilionidea, Hesperioidea und Zygaenidae: Gefährdete Arten der Schweiz, Stand 2012. Bundesamt für Umwelt BAFU, 99 pp.

Supplementary material 1

Table S1

Authors: Andreas Jaun, Hans-Peter Wymann, Kay Lucek
Data type: pdf
Explanation note: Summary of all individuals included in our study, including their sample ID, the collection date, the valley where they were collected with their respective coordinates. Individuals highlighted in bold were excluded from the genomic analyses due to their amount of missing data.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/alpento.6.80405.suppl1